Matching Schur Complement Approximations for Certain Saddle-Point Systems

John W. Pearson and Andy Wathen

[1] John W. Pearson, University of Kent, UK
: J.W.Pearson@kent.ac.uk
[2] Andy Wathen (✉), Oxford University, UK
: wathen@maths.ox.ac.uk

Summary. The solution of many practical problems described by mathematical models requires approximation methods that give rise to linear(ized) systems of equations, solving which will determine the desired approximation. This short contribution describes a particularly effective solution approach for a certain class of so-called saddle-point linear systems which arises in different contexts.

1 Introduction

Iterative methods are now widely used in various applications for the solution of linear(ized) systems of equations. A key aspect is preconditioning [38]. Without appropriate preconditioners, convergence can be unacceptably slow, whereas an effective preconditioner can *enable* the solution of matrix systems of vast dimension, and thus allow large scale computational modelling.

There continues important work on algebraic preconditioners—preconditioners which require only the entries of a (sparse) matrix for construction; triangular factorization remains an important paradigm, and algebraic multigrid techniques are finding ever wider application. However, it is now keenly realised that preconditioners which exploit matrix structures often have considerable utility. In particular, state-of-the-art preconditioners for so-called saddle-point systems [6] have found application in many areas [5, 7, 13, 15, 17, 20].

In this short contribution, we examine matrices with a particular type of saddle-point structure, namely

$$\underbrace{\begin{bmatrix} \alpha^2 A & 0 & B^T \\ 0 & \gamma^2 A & -A \\ B & -A & 0 \end{bmatrix}}_{\mathscr{A}_3} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{g} \\ \mathbf{h} \end{bmatrix}, \tag{1}$$

where $A, B \in \mathbb{R}^{n\times n}$, with A being symmetric and invertible, and with α, γ non-zero (real) parameters. We survey applications which give rise to equations of this form in section 4 below, and believe the methodology presented could be applied to application areas other than those specifically mentioned.

In fact, by simple block elimination, it is easily seen that (1) is equivalent to the 2×2 block system

$$\underbrace{\begin{bmatrix} \alpha^2 A & B^T \\ B & -\gamma^{-2} A \end{bmatrix}}_{\mathscr{A}_2} \begin{bmatrix} \mathbf{u} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{h} + \gamma^{-2}\mathbf{g} \end{bmatrix},$$

with $\mathbf{v} = \gamma^{-2}\mathbf{w} + \gamma^{-2}A^{-1}\mathbf{g}$.

Further block elimination leads to the equivalent "1×1 block system"—the Schur complement system

$$\underbrace{\left(\gamma^{-2}A + \alpha^{-2}BA^{-1}B^T\right)}_{S} \mathbf{w} = \alpha^{-2}BA^{-1}\mathbf{f} - \mathbf{h} - \gamma^{-2}\mathbf{g}.$$

In this case $\mathbf{u} = \alpha^{-2}A^{-1}\mathbf{f} - \alpha^{-2}A^{-1}B^T\mathbf{w}$, and $\mathbf{v}$ can be recovered as above.

As an alternative, one may decompose the 2×2 block system to write

$$\underbrace{\left(\alpha^2 A + \gamma^2 B^T A^{-1} B\right)}_{S_1} \mathbf{u} = \mathbf{f} + \gamma^2 B^T A^{-1}\mathbf{h} + B^T A^{-1}\mathbf{g},$$

and then recover $\mathbf{w} = \gamma^2 A^{-1}B\mathbf{u} - \gamma^2 A^{-1}\mathbf{h} - A^{-1}\mathbf{g}$ and $\mathbf{v}$ as above.

The equivalence of these 3×3, 2×2 and 1×1 block systems is well known—see, for example, [16]—and, via the result of [19, 21], the solution of any of them crucially depends on having a good approximation for the Schur complement matrix $S = \gamma^{-2}A + \alpha^{-2}BA^{-1}B^T$ or S_1. Approximations $\widehat{S}$ for which the eigenvalues of $\widehat{S}^{-1}S$ do not depend on the parameters α, γ, or on any implicit parameters (such as mesh size) which arise in A, B, are particularly valuable since they lead to iterative solvers which converge in a number of iterations independent of all such parameters, as we shall demonstrate.

Given the 3×3 block system, use of a block diagonal preconditioner

$$\mathscr{P}_3 = \begin{bmatrix} \alpha^2 A & 0 & 0 \\ 0 & \gamma^2 A & 0 \\ 0 & 0 & S \end{bmatrix}$$

allows the solution of (1) in exactly 3 iterations using the Krylov subspace iteration method MINRES [21]. Similarly, given the 2×2 block system, use of

$$\mathscr{P}_2 = \begin{bmatrix} \alpha^2 A & 0 \\ 0 & S \end{bmatrix}$$

and MINRES [22] again is guaranteed to yield the solution for any right hand side vector in 3 iterations. In either case, replacing S with a $\widehat{S}$ for which the eigenvalues of $\widehat{S}^{-1}S$ do not depend on any problem parameters yields solvers based on MINRES which require not 3, but just a few more iteration steps, and still a number independent of the parameters α, γ and the problem dimension. For the 1×1 system, the Conjugate Gradient method can also be employed effectively with $\widehat{S}$ as a preconditioner. These guarantees will be described below, but we first describe in generality a 'matching Schur complement approximation' for which the required parameter-independent eigenvalues are guaranteed.

2 Matching Schur Complement Approximation

By simple calculation it is seen that

$$S := \gamma^{-2}A + \alpha^{-2}BA^{-1}B^T = \widehat{S} - \alpha^{-1}\gamma^{-1}(B + B^T),$$

where $\widehat{S} := \left(\gamma^{-1}A + \alpha^{-1}B\right)A^{-1}\left(\gamma^{-1}A + \alpha^{-1}B\right)^T$. The original motivation for this choice of approximation, $\widehat{S}$, arose in the context of PDE-constrained optimization [27], where it was argued that the approximation allows one to match all terms except for $\alpha^{-1}\gamma^{-1}(B + B^T)$. Previous suggestions had more significant 'unmatched' terms [31].

Theorem 1. *If A is positive definite then all eigenvalues of $\widehat{S}^{-1}S$ are real, and are greater than or equal to $\frac{1}{2}$. If further the symmetric part of B is positive or negative semi-definite, then the eigenvalues of $\widehat{S}^{-1}S$ all lie in the real interval $[\frac{1}{2}, 1]$.*

Proof. The desired eigenvalues are bounded by the extreme values of the generalized Rayleigh quotient

$$R := \frac{\gamma^{-2}\mathbf{x}^T A\mathbf{x} + \alpha^{-2}\mathbf{x}^T BA^{-1}B^T\mathbf{x}}{\gamma^{-2}\mathbf{x}^T A\mathbf{x} + \alpha^{-2}\mathbf{x}^T BA^{-1}B^T\mathbf{x} + \alpha^{-1}\gamma^{-1}\mathbf{x}^T(B + B^T)\mathbf{x}}.$$

Since A is symmetric and positive definite, we can write $\mathbf{a} := \gamma^{-1}A^{1/2}\mathbf{x}$, $\mathbf{b} := \alpha^{-1}A^{-1/2}B^T\mathbf{x}$ so that

$$R = \frac{\mathbf{a}^T\mathbf{a} + \mathbf{b}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a} + \mathbf{b}^T\mathbf{b} + \mathbf{a}^T\mathbf{b} + \mathbf{b}^T\mathbf{a}} = \frac{1}{2} + \frac{1}{2}\frac{(\mathbf{a} - \mathbf{b})^T(\mathbf{a} - \mathbf{b})}{(\mathbf{a} + \mathbf{b})^T(\mathbf{a} + \mathbf{b})},$$

which evidently implies that $R \geq \frac{1}{2}$ whatever the properties of B. Further, since we are at liberty to choose the signs of α and γ, if $B + B^T$ is semi-definite then $\mathbf{a}^T\mathbf{b} + \mathbf{b}^T\mathbf{a} \geq \mathbf{0}$ with appropriate choice of signs, so that the denominator in R is clearly greater than or equal to the numerator. This gives the result. □

Some comments are in order. The multiplicative form of $\widehat{S}$ means that application of its inverse requires the solution of two systems with coefficient matrix $\gamma^{-1}A + \alpha^{-1}B$, and multiplication with A. In section 4, we describe situations where these computations are relatively straightforward using, for example, multigrid technology. That the eigenvalue spectrum is so tightly confined is somewhat remarkable, but very helpful, in particular in the context of Krylov subspace iterative methods.

Furthermore, one may use a similar analysis[3] to show that the eigenvalues of $\widehat{S}_1^{-1}S_1$ are also contained in $[\frac{1}{2}, 1]$, where

$$\widehat{S}_1 := (\alpha A + \gamma B)^T A^{-1}(\alpha A + \gamma B).$$

Both results are useful, depending on which arrangement of the saddle-point system we examine.

3 Predicted Convergence Rate of the Krylov Subspace Method

We now wish to analyze the convergence rate we can expect from an iterative method combined with our choice of preconditioner, focusing on the 3×3 block matrix $\mathscr{A}_3$ with a suitable preconditioner, applied within the MINRES algorithm.

[3] The analysis reads the same as presented for Theorem 1, except with $\mathbf{a} := \alpha A^{1/2}\mathbf{x}, \mathbf{b} := \gamma A^{-1/2}B\mathbf{x}$.

Eigenvalue bounds for preconditioned system $\widehat{\mathscr{P}}_3^{-1}\mathscr{A}_3$

Let us first consider eigenvalue bounds for $\widehat{\mathscr{P}}_3^{-1}\mathscr{A}_3$, where

$$\widehat{\mathscr{P}}_3 := \begin{bmatrix} \alpha^2\widehat{A} & 0 & 0 \\ 0 & \gamma^2\widehat{A} & 0 \\ 0 & 0 & \widehat{S} \end{bmatrix},$$

in other words where the $(1,1)$ block of the preconditioner is suitably approximated using a matrix $\widehat{A}$, and a matching strategy is used to approximate the Schur complement of $\mathscr{A}_3$.

The starting point of our analysis is the following fundamental result of Rusten and Winther [32]:

Theorem 2. *Consider the saddle-point matrix*

$$\mathscr{A}_{\Phi,\Psi} = \begin{bmatrix} \Phi & \Psi^T \\ \Psi & 0 \end{bmatrix},$$

where Φ is symmetric positive definite, and Ψ has full rank. Let $\mu_{\max}$ and $\mu_{\min}$ denote the largest and smallest eigenvalues of Φ, and let $\sigma_{\max}$ and $\sigma_{\min}$ denote the largest and smallest singular values of Ψ. Then the spectrum of $\mathscr{A}_{\Phi,\Psi}$ satisfies

$$\lambda\left(\mathscr{A}_{\Phi,\Psi}\right) \in \left[\frac{1}{2}\left(\mu_{\min} - \sqrt{\mu_{\min}^2 + 4\sigma_{\max}^2}\right), \frac{1}{2}\left(\mu_{\max} - \sqrt{\mu_{\max}^2 + 4\sigma_{\min}^2}\right)\right] \cup \left[\mu_{\min}, \frac{1}{2}\left(\mu_{\max} + \sqrt{\mu_{\max}^2 + 4\sigma_{\max}^2}\right)\right].$$

We suppose that the positive definite approximation $\widehat{A}$ is such that the eigenvalues of $\widehat{A}^{-1}A$ are contained in $[1-\zeta, 1+\eta]$, for some (preferably small) constants $\zeta \in [0,1)$, $\eta \geq 0$. Within $\widehat{\mathscr{P}}_3$, the Schur complement approximation is obtained using our matching strategy, and we assume for now that it is applied exactly.

Note that the eigenvalues of the preconditioned matrix $\widehat{\mathscr{P}}_3^{-1}\mathscr{A}_3$ are the same as those of the following (similar) matrix:

$$\begin{aligned} \widehat{\mathscr{P}}_3^{-1/2}\mathscr{A}_3\widehat{\mathscr{P}}_3^{-1/2} &= \begin{bmatrix} \alpha^2\widehat{A} & 0 & 0 \\ 0 & \gamma^2\widehat{A} & 0 \\ 0 & 0 & \widehat{S} \end{bmatrix}^{-1/2} \begin{bmatrix} \alpha^2 A & 0 & B^T \\ 0 & \gamma^2 A & -A \\ B & -A & 0 \end{bmatrix} \begin{bmatrix} \alpha^2\widehat{A} & 0 & 0 \\ 0 & \gamma^2\widehat{A} & 0 \\ 0 & 0 & \widehat{S} \end{bmatrix}^{-1/2} \\ &= \begin{bmatrix} \widehat{A}^{-1/2}A\widehat{A}^{-1/2} & 0 & \alpha^{-1}\widehat{A}^{-1/2}B^T\widehat{S}^{-1/2} \\ 0 & \widehat{A}^{-1/2}A\widehat{A}^{-1/2} & -\gamma^{-1}\widehat{A}^{-1/2}A\widehat{S}^{-1/2} \\ \alpha^{-1}\widehat{S}^{-1/2}B\widehat{A}^{-1/2} & -\gamma^{-1}\widehat{S}^{-1/2}A\widehat{A}^{-1/2} & 0 \end{bmatrix}. \end{aligned}$$

Thus consider the eigenvalues of $\widehat{\mathscr{P}}_3^{-1/2}\mathscr{A}_3\widehat{\mathscr{P}}_3^{-1/2}$. In the setting of Theorem 2,

$$\Phi = \begin{bmatrix} \widehat{A}^{-1/2}A\widehat{A}^{-1/2} & 0 \\ 0 & \widehat{A}^{-1/2}A\widehat{A}^{-1/2} \end{bmatrix}, \qquad \Psi = \begin{bmatrix} \alpha^{-1}\widehat{S}^{-1/2}B\widehat{A}^{-1/2} & -\gamma^{-1}\widehat{S}^{-1/2}A\widehat{A}^{-1/2} \end{bmatrix}.$$

First, observing that the matrix $\widehat{A}^{-1/2}A\widehat{A}^{-1/2}$ is similar to $\widehat{A}^{-1}A$ gives straightforwardly that

$$\mu_{\min} = 1-\zeta, \qquad \mu_{\max} = 1+\eta,$$

again using the notation of Theorem 2.

To find values for $\sigma_{\min}$ and $\sigma_{\max}$, we then need to look for the singular values of Ψ, which are equal to the square root of the eigenvalues of

$$\Psi\Psi^T = \alpha^{-2}\widehat{S}^{-1/2}B\widehat{A}^{-1}B^T\widehat{S}^{-1/2} + \gamma^{-2}\widehat{S}^{-1/2}A\widehat{A}^{-1}A\widehat{S}^{-1/2}. \tag{2}$$

The matrix (2) is similar to

$$\widehat{S}^{-1}\left(\alpha^{-2}B\widehat{A}^{-1}B^T + \gamma^{-2}A\widehat{A}^{-1}A\right),$$

and so its eigenvalues may be bounded by the extreme values of the Rayleigh quotient

$$\frac{\mathbf{x}^T(\alpha^{-2}B\widehat{A}^{-1}B^T + \gamma^{-2}A\widehat{A}^{-1}A)\mathbf{x}}{\mathbf{x}^T\widehat{S}\mathbf{x}} = \underbrace{\frac{\mathbf{x}^T(\alpha^{-2}B\widehat{A}^{-1}B^T + \gamma^{-2}A\widehat{A}^{-1}A)\mathbf{x}}{\mathbf{x}^T S\mathbf{x}}}_{R_1} \cdot \underbrace{\frac{\mathbf{x}^T S\mathbf{x}}{\mathbf{x}^T\widehat{S}\mathbf{x}}}_{R_2}. \tag{3}$$

Note that

$$R_1 = \frac{\mathbf{x}^T(\alpha^{-2}B\widehat{A}^{-1}B^T + \gamma^{-2}A\widehat{A}^{-1}A)\mathbf{x}}{\mathbf{x}^T(\alpha^{-2}BA^{-1}B^T + \gamma^{-2}AA^{-1}A)\mathbf{x}} = \frac{\mathbf{x}^T C\widehat{A}^{-1}C^T\mathbf{x}}{\mathbf{x}^T CA^{-1}C^T\mathbf{x}},$$

where

$$C = \begin{bmatrix} \alpha^{-1}B & \gamma^{-1}A \end{bmatrix}$$

is full rank by assumption. Thus with $\mathbf{y} = A^{-1/2}C^T\mathbf{x}$, we have

$$R_1 = \frac{\mathbf{y}^T A^{1/2}\widehat{A}^{-1}A^{1/2}\mathbf{y}}{\mathbf{y}^T\mathbf{y}},$$

from which it readily follows that $R_1 \in [1-\zeta, 1+\eta]$. We also know from Theorem 1 that $R_2 \in [\frac{1}{2}, 1]$. Putting these pieces together, we see that

$$\sigma_{\min} \geq \sqrt{\frac{1-\zeta}{2}}, \qquad \sigma_{\max} \leq \sqrt{1+\eta}.$$

Applying Theorem 2 along with our bounds for $\mu_{\min}$, $\mu_{\max}$, $\sigma_{\min}$ and $\sigma_{\max}$ gives us the following result:

Lemma 1. *The eigenvalues of the preconditioned system $\widehat{\mathscr{P}}_3^{-1}\mathscr{A}_3$ satisfy*

$$\lambda\left(\widehat{\mathscr{P}}_3^{-1}\mathscr{A}\right) \in \left[\frac{1}{2}\left(1-\zeta-\sqrt{(1-\zeta)^2+4(1+\eta)}\right), \frac{1}{2}\left(1+\eta-\sqrt{(1+\eta)^2+2(1-\zeta)}\right)\right] \cup \left[1-\zeta, \frac{1}{2}\left(1+\eta+\sqrt{5+6\eta+\eta^2}\right)\right].$$

where $\zeta \in [0,1)$ and $\eta \geq 0$ are constants such that the bounds $\lambda(\widehat{A}^{-1}A) \in [1-\zeta, 1+\eta]$ are exactly attained .

Note that in the case $\zeta = 0 = \eta$, which corresponds to the situation where the only approximation in the preconditioner $\widehat{\mathscr{P}}_3$ is the matching approximation $\widehat{S}$ for the exact Schur complement S, we have

$$\lambda(\widehat{\mathscr{P}}_3^{-1}\mathscr{A}_3) \in \left[\frac{1}{2}(1-\sqrt{5}), \frac{1}{2}(1-\sqrt{3})\right] \cup \left[1, \frac{1}{2}(1+\sqrt{5})\right].$$

Convergence rate of MINRES

It is possible to exploit the result of Lemma 1 to guarantee a resulting convergence rate of the MINRES algorithm with preconditioner $\widehat{\mathscr{P}}_3$. To do this, we make use of the following theorem [13, Theorem 4.14]:

Theorem 3. *After k steps of the preconditioned* MINRES *method, applied to a system with matrix $\mathscr{A}$ and preconditioner $\mathscr{P}$, the residual $\mathbf{r}^{(k)}$ satisfies*

$$\frac{\|\mathbf{r}^{(k)}\|_{\mathscr{P}^{-1}}}{\|\mathbf{r}^{(0)}\|_{\mathscr{P}^{-1}}} \leq 2\left(\frac{\sqrt{ad}-\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}}\right)^{\lfloor \frac{k}{2} \rfloor},$$

where $a,b,c,d > 0$ are such that $a-b=d-c$, and

$$\lambda(\mathscr{P}^{-1}\mathscr{A}) \in [-a,-b] \cup [c,d].$$

Therefore, considering our preconditioner $\widehat{\mathscr{P}}_3$ for the matrix $\mathscr{A}_3$, we may write

$$a = \frac{1}{2}\left(-1+\zeta+\sqrt{(1-\zeta)^2+4(1+\eta)}\right), \quad b = \frac{1}{2}\left(-1-\eta+\sqrt{(1+\eta)^2+2(1-\zeta)}\right),$$
$$c = 1-\zeta, \qquad\qquad d = \frac{1}{2}\left(1+\eta+\sqrt{5+6\eta+\eta^2}\right). \tag{4}$$

Clearly, in this setting, the condition $a-b=d-c$ is not satisfied. In fact it may be readily shown that $a-b<d-c$, as

$$\begin{aligned} b+d &= \frac{1}{2}\left(\sqrt{(1+\eta)^2+2(1-\zeta)}+\sqrt{(1+\eta)^2+4(1+\eta)}\right) \\ &> \frac{1}{2}\left(\sqrt{(1-\zeta)^2+2(1-\zeta)}+\sqrt{(1-\zeta)^2+4(1+\eta)}\right) \\ &> \frac{1}{2}\left(1-\zeta+\sqrt{(1-\zeta)^2+4(1+\eta)}\right) \\ &= c+a. \end{aligned}$$

However, it may clearly be stated that

$$\lambda(\widehat{\mathscr{P}}_3^{-1}\mathscr{A}_3) \in [-b+c-d,-b] \cup [c,d],$$

as the left interval has been stretched and includes the original interval $[-a,-b]$. We may use this to state the following result:

Lemma 2. *After k steps of the preconditioned* MINRES *method, applied to the 3×3 block system $\mathscr{A}_3$ and preconditioned by $\widehat{\mathscr{P}}_3$, the residual $\mathbf{r}^{(k)}$ will satisfy*

$$\frac{\|\mathbf{r}^{(k)}\|_{\widehat{\mathscr{P}}_3^{-1}}}{\|\mathbf{r}^{(0)}\|_{\widehat{\mathscr{P}}_3^{-1}}} \leq 2\left(\frac{\sqrt{d(d-c+b)}-\sqrt{bc}}{\sqrt{d(d-c+b)}+\sqrt{bc}}\right)^{\lfloor \frac{k}{2} \rfloor},$$

where b,c,d are the quantities stated in (4).

This result illustrates that the matching strategy outlined in the previous section is able to achieve rapid and robust convergence for the class of matrix systems under consideration, since the convergence bound in Lemma 2 is independent of α,γ and the dimensions of $A,\mathscr{A}$, provided only that ζ,η have such independence.

Approximate application of $\widehat{S}$

An important question is whether such a strategy can be readily applied if the Schur complement approximation $\widehat{S}$ is applied inexactly. In more detail, the matrices $L := \gamma^{-1}A + \alpha^{-1}B$ and L^T may not be straightforward to invert, so one may wish to instead approximate the Schur complement by

$$\widetilde{S} := \widehat{L}A^{-1}\widehat{L}^T,$$

where $\widehat{L}$ is some suitable (cheap) approximation of L. This then fits into the preconditioner

$$\widetilde{\mathscr{P}}_3 := \begin{bmatrix} \alpha^2\widehat{A} & 0 & 0 \\ 0 & \gamma^2\widehat{A} & 0 \\ 0 & 0 & \widetilde{S} \end{bmatrix}.$$

To analyse the performance of this preconditioner, we need to consider the eigenvalues of $\widetilde{\mathscr{P}}_3^{-1/2}\mathscr{A}_3\widetilde{\mathscr{P}}_3^{-1/2}$. Then, in the notation of Theorem 2,

$$\Phi = \begin{bmatrix} \widehat{A}^{-1/2}A\widehat{A}^{-1/2} & 0 \\ 0 & \widehat{A}^{-1/2}A\widehat{A}^{-1/2} \end{bmatrix}, \qquad \Psi = \begin{bmatrix} \alpha^{-1}\widetilde{S}^{-1/2}B\widehat{A}^{-1/2} & -\gamma^{-1}\widetilde{S}^{-1/2}A\widehat{A}^{-1/2} \end{bmatrix}.$$

The quantities $\mu_{\min}$ and $\mu_{\max}$ are therefore identical to the values when $\widehat{S}$ is applied exactly (i.e. $\widetilde{S} = \widehat{S}$) within the preconditioner.

To find suitable values for $\sigma_{\min}$ and $\sigma_{\max}$, we may apply a similar working as above, and consider the Rayleigh quotient

$$\frac{\mathbf{x}^T(\alpha^{-2}B\widehat{A}^{-1}B^T + \gamma^{-2}A\widehat{A}^{-1}A)\mathbf{x}}{\mathbf{x}^T\widetilde{S}\mathbf{x}} = \underbrace{\frac{\mathbf{x}^T(\alpha^{-2}B\widehat{A}^{-1}B^T + \gamma^{-2}A\widehat{A}^{-1}A)\mathbf{x}}{\mathbf{x}^TS\mathbf{x}}}_{R_1} \cdot \underbrace{\frac{\mathbf{x}^TS\mathbf{x}}{\mathbf{x}^T\widehat{S}\mathbf{x}}}_{R_2} \cdot \underbrace{\frac{\mathbf{x}^T\widehat{S}\mathbf{x}}{\mathbf{x}^T\widetilde{S}\mathbf{x}}}_{R_3}.$$

As for (3), we may write that $R_1 \in [1-\zeta, 1+\eta]$ and $R_2 \in [\frac{1}{2}, 1]$. We now wish to know what can be said about the quantity R_3. A useful observation, in particular if the matrix A is well conditioned, is that

$$\begin{aligned} R_3 = \frac{\mathbf{x}^TLA^{-1}L^T\mathbf{x}}{\mathbf{x}^T\widehat{L}A^{-1}\widehat{L}^T\mathbf{x}} &= \frac{\mathbf{x}^TLA^{-1}L^T\mathbf{x}}{\mathbf{x}^TLL^T\mathbf{x}} \cdot \frac{\mathbf{x}^TLL^T\mathbf{x}}{\mathbf{x}^T\widehat{L}\widehat{L}^T\mathbf{x}} \cdot \frac{\mathbf{x}^T\widehat{L}\widehat{L}^T\mathbf{x}}{\mathbf{x}^T\widehat{L}A^{-1}\widehat{L}^T\mathbf{x}} \\ &= \frac{\mathbf{y}^T\mathbf{y}}{\mathbf{y}^TA\mathbf{y}} \cdot \frac{\mathbf{x}^TLL^T\mathbf{x}}{\mathbf{x}^T\widehat{L}\widehat{L}^T\mathbf{x}} \cdot \frac{\mathbf{z}^TA\mathbf{z}}{\mathbf{z}^T\mathbf{z}}, \end{aligned}$$

where $\mathbf{y} = A^{-1/2}L^T\mathbf{x}$ and $\mathbf{z} = A^{-1/2}\widehat{L}^T\mathbf{x}$. It is clear that $\mathbf{y}^T\mathbf{y}/\mathbf{y}^TA\mathbf{y} \in [\frac{1}{\lambda_{\max}(A)}, \frac{1}{\lambda_{\min}(A)}]$ and $\mathbf{z}^TA\mathbf{z}/\mathbf{z}^T\mathbf{z} \in [\lambda_{\min}(A), \lambda_{\max}(A)]$, where $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimum and maximum eigenvalues of A, respectively.

The remaining quantity is that of $\mathbf{x}^TLL^T\mathbf{x}/\mathbf{x}^T\widehat{L}\widehat{L}^T\mathbf{x}$, and the question becomes: if $\widehat{L}$ is a good approximation of L, does this imply that $\widehat{L}\widehat{L}^T$ is a good approximation of LL^T? In general this is in fact not the case; however, as observed by Braess and Peisker in [11], if one takes $\widehat{L}$ to be m steps of a convergent iterative process applied to a symmetric L, one may state that

$$\frac{\mathbf{x}^TL^2\mathbf{x}}{\mathbf{x}^T\widehat{L}\widehat{L}^T\mathbf{x}} \in \left[(1-\omega_m)^2, (1+\omega_m)^2\right].$$

Here ω_m relates to the rate of convergence of the iterative method for L, and satisfies $\omega_m \to 0$ as $m \to \infty$. Similar observations can possibly be applied to nonsymmetric matrices L, as LL^T itself is clearly symmetric.

Using this property, we may bound the constants $\sigma_{\min}$ and $\sigma_{\max}$ as follows:

$$\sigma_{\min} \geq \sqrt{\frac{1-\zeta}{2\kappa(A)}}\,(1-\omega_m), \qquad \sigma_{\max} \leq \sqrt{(1+\eta)\kappa(A)}\,(1+\omega_m),$$

where $\kappa(A)$ denotes the condition number of A. Inserting the bounds for $\mu_{\min}$, $\mu_{\max}$, $\sigma_{\min}$ and $\sigma_{\max}$ into the result of Theorem 2 tells us that $\lambda(\widetilde{\mathscr{P}}_3^{-1}\mathscr{A}) \in [-\widetilde{a},-\widetilde{b}]\cup[c,\widetilde{d}]$, where

$$\widetilde{a} = \frac{1}{2}\left(-1+\zeta+\sqrt{(1-\zeta)^2+4(1+\eta)(1+\omega_m)^2\kappa(A)}\right),$$
$$\widetilde{b} = \frac{1}{2}\left(-1-\eta+\sqrt{(1+\eta)^2+\frac{2(1-\zeta)}{\kappa(A)}(1-\omega_m)^2}\right),$$
$$\widetilde{d} = \frac{1}{2}\left(1+\eta+\sqrt{(1+\eta)^2+4(1+\eta)(1+\omega_m)^2\kappa(A)}\right).$$

We note that these bounds for the eigenvalues of $\widetilde{\mathscr{P}}_3^{-1}\mathscr{A}$ are weak, sometimes extremely so, as when the approximations of L become increasingly accurate (i.e. $\omega_m \to 0$), the values of $\widetilde{a},\widetilde{b},\widetilde{d}$ should tend to those of a,b,d in (4). However, in the above expressions, the factors of $\kappa(A)$ remain when inserting $\omega_m = 0$. Therefore, if $\kappa(A)$ is well conditioned, as for many problems in PDE-constrained optimization for example, the theoretical guarantee of the effectiveness of $\widetilde{\mathscr{P}}_3$ is obtained straightforwardly. If this is not the case, this highlights the necessity of a potent scheme to approximate the (inverse action of) L and L^T appropriately. In practice, a number of cycles of a tailored multigrid scheme is often found to perform this function, for instance.

We highlight that the theoretical issues surrounding the approximation of matrices of the form LL^T is not restricted to the matching strategy presented in this paper, and arises when using many different preconditioners for saddle-point systems, due to the structure of the Schur complement of the saddle-point system itself.

Comments on 2×2 and 1×1 block cases

The analysis presented in this section has focused on solving matrix systems of the structure $\mathscr{A}_3$ using preconditioner $\widehat{\mathscr{P}}_3$ within the MINRES algorithm. Of course, it is perfectly legitimate to reduce the system to the form $\mathscr{A}_2$, and solve this using MINRES with preconditioner

$$\widehat{\mathscr{P}}_2 := \begin{bmatrix} \alpha^2\widehat{A} & 0 \\ 0 & \widehat{S} \end{bmatrix}.$$

Literature such as [33] considers eigenvalue bounds for saddle-point systems with non-zero $(2,2)$ block, which arise when considering the matrix of importance in this case:

$$\widehat{\mathscr{P}}_2^{-1/2}\mathscr{A}_2\widehat{\mathscr{P}}_2^{-1/2} = \begin{bmatrix} \widehat{A}^{-1/2}A\widehat{A}^{-1/2} & \alpha^{-1}\widehat{A}^{-1/2}B^T\widehat{S}^{-1/2} \\ \alpha^{-1}\widehat{S}^{-1/2}B\widehat{A}^{-1/2} & -\gamma^{-2}\widehat{S}^{-1/2}A\widehat{S}^{-1/2} \end{bmatrix}.$$

The analysis in this case of 2×2 blocks is more standard and is summarised, for example in Chapter 4 of [13], or [28]. It can be applied to demonstrate that a similar MINRES convergence rate to the 3×3 case is to be expected for such 2×2 systems when the same approximations

are employed within the preconditioner. The matching strategy is also, therefore, an effective approach for systems of the form $\mathscr{A}_2$.

It is also possible to consider the Schur complement (1×1 block) system itself, and apply preconditioned Conjugate Gradients with our matching strategy. In this case the potency of the iterative method will depend directly on the effectiveness of the matching strategy, which we have ascertained to guarantee compact eigenvalue bounds. However, we emphasize that such a solver will require a matrix-vector multiplication with S or S_1, and therefore an exact representation of A^{-1} will generally be required. Such a method should therefore only be applied if A has a simple structure, for instance if it is a diagonal matrix.

4 Applications of Matching Approach

In this section, we wish to briefly survey applications in which the matching strategy discussed in this paper has been applied.

PDE-constrained optimization: The class of problems for which the authors originally derived this approach was that of PDE-constrained optimization problems of the following form:

$$\begin{aligned} \min_{y,c} \quad & \frac{1}{2}\|y-\widehat{y}\|^2_{L_2(\Omega)} + \frac{\beta}{2}\|c\|^2_{L_2(\Omega)} \\ \text{s.t.} \quad & \mathscr{L}y = c, \quad \text{in } \Omega, \\ & y = h, \quad \text{on } \partial\Omega. \end{aligned}$$

Here, y and c denote *state* and *control variables* which we wish to find, with $\widehat{y}$ a given *desired state* and $\beta > 0$ a *regularization parameter*. The constraints of the optimization problem are derived from a PDE operator $\mathscr{L}$ on a domain Ω, and given Dirichlet boundary conditions h on the boundary $\partial\Omega$ of the domain. Other boundary conditions are possible, though boundary control problems have a slightly different form [18].

If the PDE operator $\mathscr{L} = -\nabla^2 + \mathbf{w}\cdot\nabla$, where $\mathbf{w}$ is some given wind vector, then the problem under consideration is that of *convection–diffusion control*, and upon discretization of this problem the matrix system to be solved is [26, 27]

$$\begin{bmatrix} M & 0 & \bar{K}^T \\ 0 & \beta M & -M \\ \bar{K} & -M & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{c} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \\ \mathbf{h} \end{bmatrix}, \tag{5}$$

with $\mathbf{y}$, $\mathbf{c}$ and $\mathbf{p}$ the discretized state, control and *adjoint variables*, and $\mathbf{f}$ and $\mathbf{h}$ including terms arising from the desired state and boundary conditions. Here, M is a finite element mass matrix which is symmetric positive definite, and $\bar{K}$ is a finite element matrix relating to $\mathscr{L}$ which has the property that $\bar{K}+\bar{K}^T$ is positive semidefinite. If $\mathbf{w} = \mathbf{0}$, then the control problem reduces to that of *Poisson control*, and $\bar{K} = \bar{K}^T = K$ is a stiffness matrix. For either problem, the system (5) is of the form $\mathscr{A}_3$, with $A = M$, $B = \bar{K}$, $\alpha = 1$, $\gamma = \sqrt{\beta}$, and the theory of this paper can be applied (see [26, 27]). Such problems have the additional convenient property that the mass matrix M is well conditioned.

This theory has been extended to a range of other PDE-constrained optimization problems of different structure, for example to time-independent and time-dependent fluid flow control problems [23, 24, 37], reaction-diffusion control problems from chemical reactions and pattern formation processes in mathematical biology [25, 36], and active set Newton methods for

problems with additional bound constraints [29]. Further, the papers [3, 4] examine PDE-constrained optimization problems with uncertain inputs using this strategy, low-rank methods are derived for a class of time-dependent problems in [35], and optimization problems with fractional differential equation constraints are studied in [12] (where a result of the type shown in Theorem 1 is proved using the Binomial Theorem).

Complex valued linear systems: Matrix systems of similar type to the 2×2 block form $\mathscr{A}_2$ are discussed in [1] in the context of complex valued linear systems. In more detail, consider the solution of the complex matrix system $C\mathbf{z} = \mathbf{d}$, where $C = A + iB$, $\mathbf{z} = \mathbf{u} + i\mathbf{w}$ and $\mathbf{d} = \mathbf{f} + i\mathbf{h}$. Therefore $(A + iB)(\mathbf{u} + i\mathbf{w}) = \mathbf{f} + i\mathbf{h}$, whereupon comparing real and imaginary parts gives the matrix system

$$\begin{bmatrix} A & -B \\ B & A \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{h} \end{bmatrix}. \tag{6}$$

It is clear that, if A and B are symmetric, one may rearrange the system (6) to a symmetric matrix of form $\mathscr{A}_2$, with $\alpha = \gamma = 1$.

In [1], the authors derive a preconditioner for the system (6) based on the matching strategy. Further, in [2], preconditioned modified Hermitian and skew-Hermitian splitting (PMHSS) iteration methods for 2×2 block linear systems are considered using the same methodology.

Cahn–Hilliard models: Another major application area of the approach we have outlined is that of the numerical solution of Cahn–Hilliard models describing phase separation. For instance, in [9] the authors consider the H^{-1}–gradient flow of the Ginzburg–Landau energy

$$E(u) := \int_\Omega \frac{\delta\varepsilon}{2} |\nabla u|^2 + \frac{1}{\varepsilon}\psi(u)\, \mathrm{d}\Omega,$$

with $\delta, \varepsilon > 0$, and an *obstacle potential* given by $\psi(u) = \frac{1}{2}(1-u^2) + I_{[-1,1]}(u)$ with an indicator function I (though there are other possible choices for this potential).

Upon discretizing the resulting PDEs, the authors are required to solve matrix systems of the form

$$\begin{bmatrix} -H & M \\ M & \tau K \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{h} \end{bmatrix},$$

where M, K are defined as above, H is a symmetric matrix which involves the sum of a stiffness matrix and terms involving $\psi'(u)$, and $\tau > 0$ is the time-step used within the numerical method. Although this system is not precisely of the form $\mathscr{A}_2$, the authors were able to use convenient properties of H to obtain good numerical results for certain restrictions of the time-step (i.e. $\tau < \varepsilon^2$). In [8] some theoretical guarantees are provided for similar preconditioners for image inpainting problems.

We note that many scientists have applied the matching strategy to Cahn–Hilliard models. In [10] preconditioners for large scale binary Cahn–Hilliard models are considered, matrix systems arising from the evolution of diblock copolymer melts are tackled in [14], and solvers for the phase field crystal equation, which is itself of Cahn–Hilliard type, are constructed in [30].

We note that the fields categorised above do not represent an exhaustive list of applications for the approach presented in this paper. For instance, see [34] for a discussion of preconditioners for discontinuous Galerkin time-stepping methods, and many other papers discussing optimal control problems and Cahn–Hilliard equations, for other recent developments of this method.

5 Concluding Remarks

We have considered block preconditioners for a particular class of saddle-point matrices which arise in various applications. Specifically, we have demonstrated the efficacy of an approach which employs a 'matching strategy' for the approximation of a Schur complement. The use of the resulting preconditioners is shown to enable the iterative solution of corresponding systems of equations in a number of iterations independent of parameters in the problem and of the dimension of the relevant matrices. This is therefore a highly effective solution approach for such systems of equations.

Acknowledgement. Ian Sloan has been a great friend and mentor over many years. Through his mathematical insight, he has derived really useful approximations in different contexts throughout his long career. We wish to express our deep appreciation for his leadership in our field and to acknowledge his encouragement of all those who seek to contribute to it.

References

1. Axelsson, O., Neytcheva, M., Ahmad, B.: A comparison of iterative methods to solve complex valued linear algebraic systems, Numer. Algor. **66**, 811–841 (2014).
2. Bai, Z.-Z., Benzi, M., Chen, F., Wang, Z.-Q.: Preconditioned MHSS iteration methods for a class of block two-by-two linear systems with applications to distributed control problems, IMA J. Numer. Anal. **33**, 343–369 (2013).
3. Benner, P., Dolgov, S., Onwunta, A., Stoll, M.: Low-rank solvers for unsteady Stokes-Brinkman optimal control problem with random data, Comput. Methods Appl. Mech. Eng. **304**, 26–54 (2016).
4. Benner, P., Onwunta, A., Stoll, M.: Block-diagonal preconditioning for optimal control problems constrained by PDEs with uncertain inputs, SIAM J. Matrix Anal. Appl. **37**(2), 491–518 (2016).
5. Benzi, M., Haber, E., Taralli, L.: Multilevel algorithms for large-scale interior point methods, SIAM J. Sci. Comput. **31**(6), 4152–4175 (2009).
6. Benzi, M., Golub, G.H., Liesen, J.: Numerical solution of saddle point problems, Acta Numerica **14**, 1–137 (2005).
7. Biros, G., Ghattas, O.: Parallel Lagrange–Newton–Krylov–Schur methods for PDE-constrained optimization. Part I: The Krylov–Schur solver, SIAM J. Sci. Comput. **27**(2), 687–713 (2005).
8. Bosch, J., Kay, D., Stoll, M., Wathen, A.J.: Fast solvers for Cahn–Hilliard inpainting, SIAM J. Imaging Sci. **7**(1), 67–97 (2014).
9. Bosch, J., Stoll, M., Benner, P.: Fast solution of Cahn-Hilliard variational inequalities using implicit time discretization and finite elements, J. Comput. Phys. **262**, 38–57 (2014).
10. Boyanova, P., Do-Quang, M., Neytcheva, M: Efficient preconditioners for large scale binary Cahn–Hilliard models, Comput. Methods Appl. Math. **12**(1), 1–22 (2012).
11. Braess, D., Peisker, D.: On the numerical solution of the biharmonic equation and the role of squaring matrices for preconditioning, IMA J. Numer. Anal. **6**, 393–404 (1986).
12. Dolgov, S., Pearson, J.W., Savostyanov, D.V., Stoll, M.: Fast tensor product solvers for optimization problems with fractional differential equations as constraints, Appl. Math. Comput. **273**, 604–623 (2016).

13. Elman, H.C., Silvester, D.J., Wathen, A.J.: Finite Elements and Fast Iterative Solvers: With Applications in Incompressible Fluid Dynamics, Oxford University Press, New York (2014).
14. Farrell, P.E., Pearson, J.W.: A preconditioner for the Ohta–Kawasaki equation, submitted (2016).
15. Gill, P.E., Murray, W., Wright, M.H.: Practical Optimization, Academic Press (1982).
16. Greif, C., Moulding, E., Orban, D.: Bounds on eigenvalues of matrices arising from interior-point methods, SIAM J. Optim. **24**(1), 49–83 (2014).
17. Haber, E., Ascher, U.M.: Preconditioned all-at-once methods for large, sparse parameter estimation problems, Inverse Problems **17**(6), 1847–1864 (2001).
18. Heidel, G., Wathen, A.J.: Preconditioning for boundary control problems in incompressible fluid dynamics, submitted to SIAM J. Sci. Comput. (2015).
19. Ipsen, I.C.F.: A note on preconditioning nonsymmetric matrices, SIAM J. Sci. Comput. **23**(3), 1050–1051 (2001).
20. Le Gia, Q.T., Sloan, I.H., Wathen, A.J.: Stability and preconditioning for a hybrid approximation on the sphere, Numer. Math. **118**(4), 695–711 (2011).
21. Murphy, M.F., Golub, G.H., Wathen, A.J.: A note on preconditioning for indefinite linear systems, SIAM J. Sci. Comput. **21**(6), 1969–1972 (2000).
22. Paige, C.C., Saunders, M.A.: Solution of sparse indefinite systems of linear equations, SIAM J. Numer. Anal., **12**, 617–629 (1975).
23. Pearson, J.W.: Preconditioned iterative methods for Navier–Stokes control problems, J. Comput. Phys. **292**, 194-207 (2015).
24. Pearson, J.W.: On the development of parameter-robust preconditioners and commutator arguments for solving Stokes control problems, Electron. Trans. Numer. Anal. **44**, 53–72 (2015).
25. Pearson, J.W., Stoll, M.: Fast iterative solution of reaction-diffusion control problems arising from chemical processes, SIAM J. Sci. Comput. **35**(5), B987–B1009 (2013).
26. Pearson, J.W., Wathen, A.J.: Fast iterative solvers for convection-diffusion control problems, Electron. Trans. Numer. Anal. **40**, 294–310 (2013).
27. Pearson, J.W., Wathen, A.J.: A new approximation of the Schur complement in preconditioners for PDE-constrained optimization, Numer. Linear Alg. Appl. **19**(5), 816–829 (2012).
28. Pestana, J., Wathen, A.J.: Natural preconditioning and iterative methods for saddle point systems, SIAM Rev. **57**(1), 71-91 (2015).
29. Porcelli, M., Simoncini, V., Tani, M.: Preconditioning of active-set Newton methods for PDE-constrained optimal control problems, SIAM J. Sci. Comput. **37**(5), S472–S502 (2014).
30. Praetorius, S., Voigt, M.: Development and analysis of a block-preconditioner for the phase-field crystal equation, SIAM J. Sci. Comput. **37**(3), B425–B451 (2015).
31. Rees, T., Dollar, H.S., Wathen, A.J.: Optimal solvers for PDE-constrained optimization, SIAM J. Sci. Comput. **32**(1), 271–298 (2010).
32. Rusten, T., Winther, R.: A preconditioned iterative method for saddle point problems, SIAM J. Matrix Anal. Appl. **13**(3), 887–904 (1992).
33. Silvester, D., Wathen, A.: Fast iterative solution of stabilised Stokes systems. Part II: using general block preconditioners, SIAM J. Numer. Anal. **31**(5), 1352–1367 (1994).
34. Smears, I.: Robust and efficient preconditioners for the discontinuous Galerkin time-stepping method, to appear in IMA J. Numer. Anal. (2016).
35. Stoll, M., Breiten, T.: A low-rank in time approach to PDE-constrained optimization, SIAM J. Sci. Comput. **37**(1), B1–B29 (2015).

36. Stoll, M., Pearson, J.W., Maini, P.K.: Fast solvers for optimal control problems from pattern formation, J. Comput. Phys. **304**, 27–45 (2016).
37. Stoll, M., Wathen, A.: All-at-once solution of time-dependent Stokes control, J. Comput. Phys. **232**(1), 498–515 (2013).
38. Wathen, A.J.: Preconditioning, Acta Numerica **24**, 329–376 (2015).